

**TOWARDS UNBIASED ESTIMATION
OF THE OPTIMAL TIME FOR RENAL REPLACEMENT THERAPY**

by

Florentina Elena Sileanu

BS, University of Pittsburgh, 2009

Submitted to the Graduate Faculty of
the Graduate School of Public Health in partial fulfillment
of the requirements for the degree of
Master of Science

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH
GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Florentina Elena Sileanu

It was defended on

April 22, 2015

and approved by

Thesis Advisor:

Chung-Chou H. Chang, PhD
Professor, Departments of Medicine and Biostatistics
School of Medicine and Graduate School of Public Health
University of Pittsburgh

Committee Members:

Gilles Clermont, MD, MSc
Professor, Department of Critical Care Medicine
School of Medicine
University of Pittsburgh

John A. Kellum, MD, MCCM
Professor, Department of Critical Care Medicine
School of Medicine
University of Pittsburgh

Ada Youk, PhD
Associate Professor, Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Copyright © by Florentina Elena Sileanu

2015

**TOWARDS UNBIASED ESTIMATION
OF THE OPTIMAL TIME FOR RENAL REPLACEMENT THERAPY**

Florentina Elena Sileanu, MS

University of Pittsburgh, 2015

ABSTRACT

Acute kidney injury (AKI), a sudden deterioration in renal function which occurs when the kidneys no longer remove waste products from the blood, is a challenging medical condition that affects intensive care unit patients worldwide. Patients with severe cases of AKI are placed on renal replacement therapy (RRT), a life-supporting treatment, and have been linked to mortality rates as high as 60%. Despite having guidelines with indications for RRT it is unclear what the optimal initiation time should be. Studies looking at the association between timing of initiation and mortality give contradictory results: some suggest a better outcome with early initiation while others with late initiation. There are four issues with current studies: 1) selection bias due to treatment status being driven by a patient's baseline characteristics and the physician's decision to treat; 2) the time from which survival is measured is different across studies causing lead-time or immortal-time biases; 3) results from the different statistical methods used are not always comparable; 4) patients never started on RRT are excluded from analyses.

The aim of this study is to determine the association between timing of initiation of RRT and mortality by addressing existing biases and limitations. Selection bias will be controlled for by a propensity score and 1-1 matching without replacement using the nearest neighbor Mahalanobis distance. Lead-time bias will be addressed by counting survival time from the

same point for all patients. Immortal-time bias will be eliminated by using an expanded risk sets analysis in which patients are part of all three risk groups: early, late, and no RRT. Unlike current studies patients never started on RRT will also be analyzed. Cox proportional hazards will be used to test differences in the hazard of mortality at 1-year between groups.

Public Health Significance: To our knowledge, this is the largest observational study investigating the optimal time for initiating RRT. Our study shows the effect of different biases on the outcome and reinforces the importance of carefully designing an observational study. Future nephrology researchers can use this work as foundation in the quest of finding the optimal time for RRT initiation.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	X
1.0 INTRODUCTION.....	1
1.1 SIGNIFICANCE	2
2.0 MATERIALS AND METHODS	5
2.1 STATISTICAL CONCEPTS	5
2.1.1 Propensity Score in Observational Studies.....	5
2.1.2 Mahalanobis Distance Matching.....	6
2.1.3 Cox Proportional Hazards Model	8
2.1.4 Expanded Risk Sets.....	9
2.2 TIMING OF RRT STUDY DATA	11
2.2.1 Study Population.....	11
2.2.2 Data Collection	12
2.2.3 Timing of RRT	13
2.2.4 Outcome Assessment.....	13
2.2.5 Statistical Analyses.....	14
3.0 RESULTS.....	15
3.1 BASELINE CHARACTERISTICS.....	15

3.2	MATCHING.....	17
3.2.1	Generating Propensity Scores	17
3.2.2	Generating Matched Populations	18
3.2.3	Generating Expanded Risk Sets Populations	23
3.3	SURVIVAL.....	23
3.3.1	Before Matching.....	23
3.3.2	After Matching.....	25
4.0	DISCUSSION	28
	BIBLIOGRAPHY.....	32

LIST OF TABLES

Table 1: Summary of studies evaluating the timing of initiation of RRT	2
Table 2: Patient characteristics by in-hospital RRT.....	16
Table 3: Sample size description by early RRT definition and analysis type	18
Table 4: Unadjusted outcomes before matching	24
Table 5: Survival at 1-year in the matched and expanded risk sets populations	26

LIST OF FIGURES

Figure 1: Illustration of lead-time bias and immortal time bias	4
Figure 2: Illustration of Euclidean distance (a) and Mahalanobis distance (b)	7
Figure 3: Follow-up of subjects from different RRT groups in the expanded risk sets	10
Figure 4: Study population	12
Figure 5: Jitter plot of propensity scores in the matched and unmatched groups	21
Figure 6: Covariate balance before and after matching	22
Figure 7: Age adjusted survival at 1-year before matching	25
Figure 8: Age adjusted survival at 1-year after matching	27

ACKNOWLEDGMENTS

This research could not have been conducted without the close mentoring of Doctors Gilles Clermont, John Kellum and Raghavan Murugan. The statistical accuracy of the methods was reviewed by Doctors Chung-Chou H. Chang and Ada Youk.

I thank you all for your support and continued mentorship.

1.0 INTRODUCTION

Acute kidney injury (AKI) is a challenging condition characterized by an abrupt decline in kidney function over a period of hours to days that can occur before or in the hospital setting.^{1,2} Worldwide, severe AKI occurs in approximately 6% of intensive care unit (ICU) patients, with almost two-thirds receiving renal replacement therapy (RRT).³ For severe AKI patients hospital mortality is approximately 60% and dialysis dependence at hospital discharge is approximately 14%.³ AKI has also been associated with increased length of ICU and hospital stay, with those in need of RRT having a median ICU stay 3 times longer than those without AKI.¹

The association between timing of initiation of RRT and mortality is uncertain. The Kidney Disease Improving Global Outcomes (KDIGO) Clinical Practice Guidelines recommend starting RRT based on the clinical context, the presence of conditions that can be modified with RRT, and trends of laboratory tests rather than single blood urea nitrogen or creatinine levels.² Despite these guidelines, hard data remain absent or conflictive regarding the optimal time to start dialysis.⁴ Some studies suggest that early initiation of RRT is associated with lower mortality⁵⁻⁸, other studies suggest no difference^{9,10}, while a recent multicenter retrospective observational study found a U-shape association between RRT timing and in-hospital mortality.¹¹ **(Table 1)**

Table 1: Summary of studies evaluating the timing of initiation of RRT

Study	Year	Study Design	N	BUN at initiation of RRT (mg/dl)		Hospital Mortality (%)	
				Early	Late	Early	Late
Gettings <i>et al.</i> ⁵	1999	Retrospective	100	<60	≥60	61	80
Demirkiliç <i>et al.</i> ⁶	2004	Retrospective ^a	61	NS	NS	23.5	55.5
Elahi <i>et al.</i> ⁷	2004	Retrospective ^a	64	67±35 ^b	75±61 ^b	22	43
Liu <i>et al.</i> ⁸	2006	Observational	243	≤76	76	35 ^c	41 ^c
Korevaar <i>et al.</i> ⁹	2001	Prospective ^d	253	NS	NS	16 ^e	25 ^e
Bouman <i>et al.</i> ¹⁰	2002	RCT	106	LV: 48 (40-66) ^f HV: 46 (38-58) ^f	LV: 105 (62-116) ^f	LV: 51 HV: 37	LV: 40
Shiao <i>et al.</i> ¹¹	2012	Retrospective ^g	648	EG: 49±29 ^b IG: 59±26 ^b	LG: 90±43 ^b	EG: 59 IG: 49	LG: 67

^aRRT started based on urine output <100 ml over 8 hours in early group and based on biochemical parameters in late group. ^bMean BUN ± standard deviation. ^cPercent of patients that died by day 28 from ICU admission. ^dClassification into early and late was done according to the Dialysis Outcomes Quality Initiative¹². ^ePercent of patients that died during the 24 months after RRT initiation. ^fMedian BUN (BUN quartiles). ^gEG, early group, ≤1 day; IG, intermediate group, 2–3 days; LG, late group, ≥4 days between ICU admission and RRT initiation. BUN, serum blood urea nitrogen; RRT, renal replacement therapy; NS, not specified; RCT, randomized controlled trial; LV, low-volume hemofiltration; HV, high-volume hemofiltration; ICU, intensive care unit.

1.1 SIGNIFICANCE

There are four main shortcomings in current studies that attempt to define the optimal time for initiating RRT. First, there are absolute and relative indications for RRT initiation and the decision to start therapy is affected by strongly held physician beliefs, patient characteristics, and the logistical or organizational aspects of a given institution.¹³ The proportion of patients with absolute or relative indications may vary across studies leading to selection bias. Second,

the time from which survival is measured is not defined the same in all studies. Some studies measured survival from RRT initiation^{9,11}, while others measured survival from a fixed time point prior to RRT initiation^{5-8,10} which causes lead-time bias and immortal-time bias respectively. Suppose patient A starts RRT on the same day as KDIGO stage 3 (*i.e.* the baseline time point from which survival is measured), patient B starts RRT a few days after baseline and both patients are followed from their RRT initiation until death or censoring. In this case, patient A has an artificial survival advantage, or lead-time bias, since at the time of RRT initiation he/she was earlier in the course of disease progression than patient B (**Figure 1** top). If both patients are followed from baseline until death or censoring patient B has an artificial survival advantage, or immortal-time bias, because he/she had to survive between baseline and RRT initiation (**Figure 1** bottom). Third, the statistical methods used are different between studies and their results are not always directly comparable. In some studies the main outcome was the crude hospital mortality rate⁵⁻⁷ while other studies looked at time to event analyses.⁸⁻¹¹ Fourth, the exclusion from all current studies of patients that were never started on RRT due to recovery, death or lost to follow-up severely limits their validity.¹⁴

The aim of this study is to determine the association between timing of initiation of RRT and mortality by addressing the existing biases and limitations in the current literature. Treatment selection bias will be controlled for by the use of a propensity score¹⁵ and lead-time and immortal-time biases will be addressed by using an expanded risk sets (ERS) analysis.¹⁶ Unlike current studies we will also take into account patients that were never started on RRT.

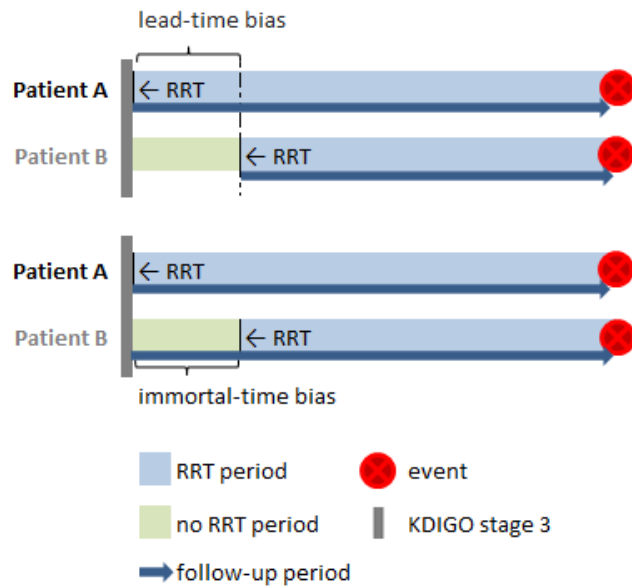


Figure 1: Illustration of lead-time bias and immortal-time bias
 RRT, renal replacement therapy; KDIGO stage 3 is the baseline time from which survival is measured.

2.0 MATERIALS AND METHODS

2.1 STATISTICAL CONCEPTS

2.1.1 Propensity Score in Observational Studies

Estimating treatment effects in observational studies suffers from unmeasured confounding and selection bias due to treatment status being driven by a patient's baseline characteristics and the physician's decision to treat. As a result, baseline characteristics differ systematically between treated and untreated subjects. Propensity score matching (PSM) methods are recommended in order to adjust for such unmeasured confounding and selection bias.¹⁷

Per Rosenbaum and Rubin (1983) the propensity score is the probability towards treatment assignment conditional on observed baseline covariates. In practice, the propensity score is usually estimated by logistic regression:

$$\hat{\pi}_i = \hat{E}(T_i | X_i) = \hat{P}(T_i = 1 | X_i) = \frac{e^{X_i \hat{\beta}}}{1 + e^{X_i \hat{\beta}}}.$$

As such, the propensity score is a balancing score that allows the selection of treated ($t = 1$) and control ($t = 0$) subjects with similar distributions of observed baseline covariates, making the two groups directly comparable.

2.1.2 Mahalanobis Distance Matching

Matching on the propensity score and any function of the observed baseline covariates will also balance the treatment and control groups.¹⁸ An implementation of this is the nearest available Mahalanobis distance which takes into account the variance and the covariance between all variables used in the calculation of distance.

Let \vec{x}_i be a vector of p observed baseline covariates for subject i from the treatment group and \vec{x}_j be a vector of p observed baseline covariates for subject j from the control group. Let Σ be the sample variance-covariance matrix, defined as:

$$\Sigma_{x_i, x_j} = \begin{bmatrix} \text{var}(x_i) & \text{cov}(x_i, x_j) \\ \text{cov}(x_i, x_j) & \text{var}(x_j) \end{bmatrix}.$$

Then, the Mahalanobis distance can be defined as:

$$M_d(i, j) = \sqrt{(\vec{x}_i - \vec{x}_j)^T \Sigma_{x_i, x_j}^{-1} (\vec{x}_i - \vec{x}_j)}.$$

For uncorrelated variables with unit variance $M_d(i, j)$ reduces to the Euclidean distance. **Figure 2** provides an illustration of the Euclidean distance (a) and the Mahalanobis distance (b) where the contours represent equidistant points from the center using each distance metric.¹⁹ It can be seen that the Euclidean distance treats the data as if it had a spherical distribution while the Mahalanobis distance takes into account the distribution of the data points.

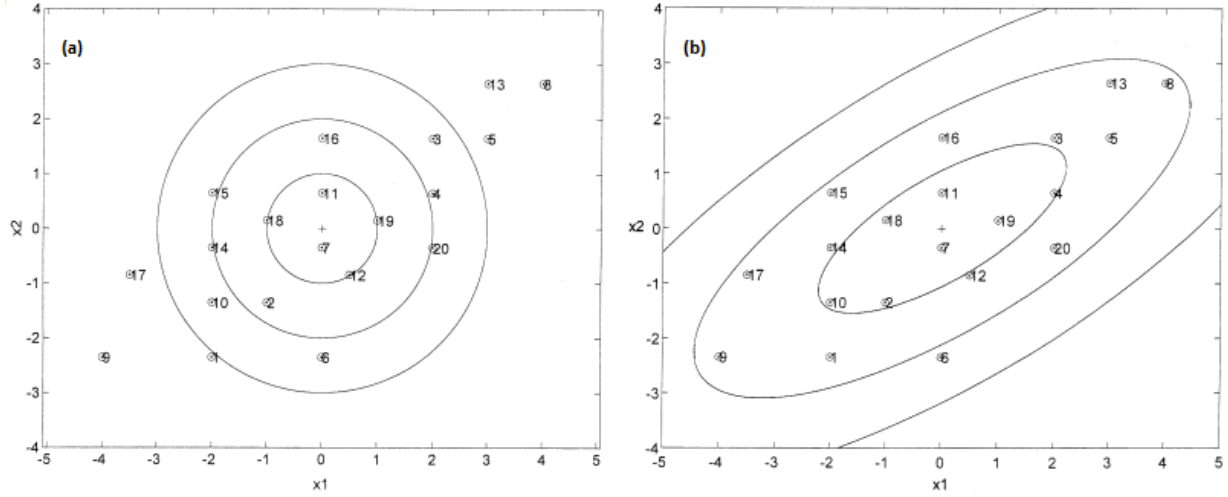


Figure 2: Illustration of Euclidean distance (a) and Mahalanobis distance (b)

Circles represent equal Euclidean distances towards the center point and ellipses represent equal Mahalanobis distances towards the center point.¹⁹

Assuming that the propensity score has been generated, in order to calculate M_d and create a matched dataset, the following steps have to be taken:

1. Transform the raw data X through spectral decomposition of Σ into X^* which has an identity covariance;
2. Randomly order the subjects from the treatment group and those from the control group;
3. Calculate all pairwise Mahalanobis distances based on X^* ;
4. Choose the first subject i from the treatment group and find subject j from the control group that has the smallest $M_d(i, j)$;
5. Remove pair (i, j) from the pool, move to the second subject from the treatment group and apply steps 4 and 5 until there are no more subjects in the treatment group.

2.1.3 Cox Proportional Hazards Model

For each i -th subject ($i=1, \dots, n$) let $T_i = \min(T_i^*, C_i)$ be the observed follow-up time given by the minimum between the event time T_i^* and the censoring time C_i , $\delta_i = I(T_i^* \leq C_i)$ the event indicator which is 1 when $T_i^* \leq C_i$ and 0 otherwise, and X_i a vector of p baseline covariates. We are interested in estimating the marginal survival function at time t given by $S(t) = P(T > t)$ and adjusted for covariates X where the observed right-censored survival data will be represented by $\{T_i, \delta_i, X_i\}$ for each of the n subjects. In such a setting, Cox proportional hazards (PH) model is typically used.²⁰ Cox PH is defined as:

$$h(t | X) = h_0(t) e^{X' \beta},$$

where $h_0(t)$ is the unknown baseline hazard function for $X = 0$ and β a p -dimensional vector of unknown parameters. The estimated marginal survival function under the Cox PH model is defined as:

$$\hat{S}(t | X) = e^{-\hat{H}_0(t) e^{X' \hat{\beta}}},$$

where $\hat{\beta}$ is the maximum-likelihood estimator of β and $\hat{H}_0(t)$ is the Breslow estimator of $H_0(t)$. This model is subject to the proportionality assumption which implies that the survival curves for different X strata must have hazard functions that are proportional over time.

2.1.4 Expanded Risk Sets

In order for the start of follow-up for survival to be the same for all RRT groups one would have to assign subjects to their RRT group on the day of KDIGO stage 3 which is impossible without knowing the future treatment path for each subject. Suppose early RRT was defined as having started RRT within 3 days of KDIGO stage 3 and subject A started RRT on day 2 and died on day 19, subject B started RRT on day 4 and was alive at one year, subject C never started RRT and died on day 4 and subject D never started RRT and was alive at one year. In retrospect we would say that subject A was an early starter, subject B a late starter and subjects C and D were part of the no RRT group. However, following the subjects prospectively it is not until days 2, 4, 4, and one year that we know the true groups for subjects A, B, C, and D respectively. Thus the true RRT group is determined after the start of follow-up and it is contrived to assign subjects to a group on the day of KDIGO stage 3.

Through the use of ERS analysis subjects are followed prospectively from the time they reach KDIGO stage 3 and are allowed to have different contributions of follow-up times and events in all three risk sets: early RRT, late RRT, and no RRT. **Figure 3** displays the ERS replicate contributions for subjects A, B, C, and D. For example, subject A was followed in the early RRT group for 19 days and contributed with an event to this group. Subject A also had the potential of being part of the late RRT and no RRT groups for 2 days until he/she was artificially censored because RRT was initiated. Subject B had the potential to be part of the early RRT group for 3 days and was censored, was followed in the late RRT group for one year and was censored, and had the potential to be part of the no RRT group for 4 days, until RRT was initiated, and as a

result became censored. Subject C had the potential to be part of the early RRT group for 3 days and was censored, was followed in the late RRT and no RRT groups for 4 days, until his/her death. Thus he/she contributed with an event to both late and no RRT groups. Subject D had the potential to be part of the early RRT group for 3 days, was followed in the late RRT and no RRT groups for one year and did not contribute with an event to any of the three groups.

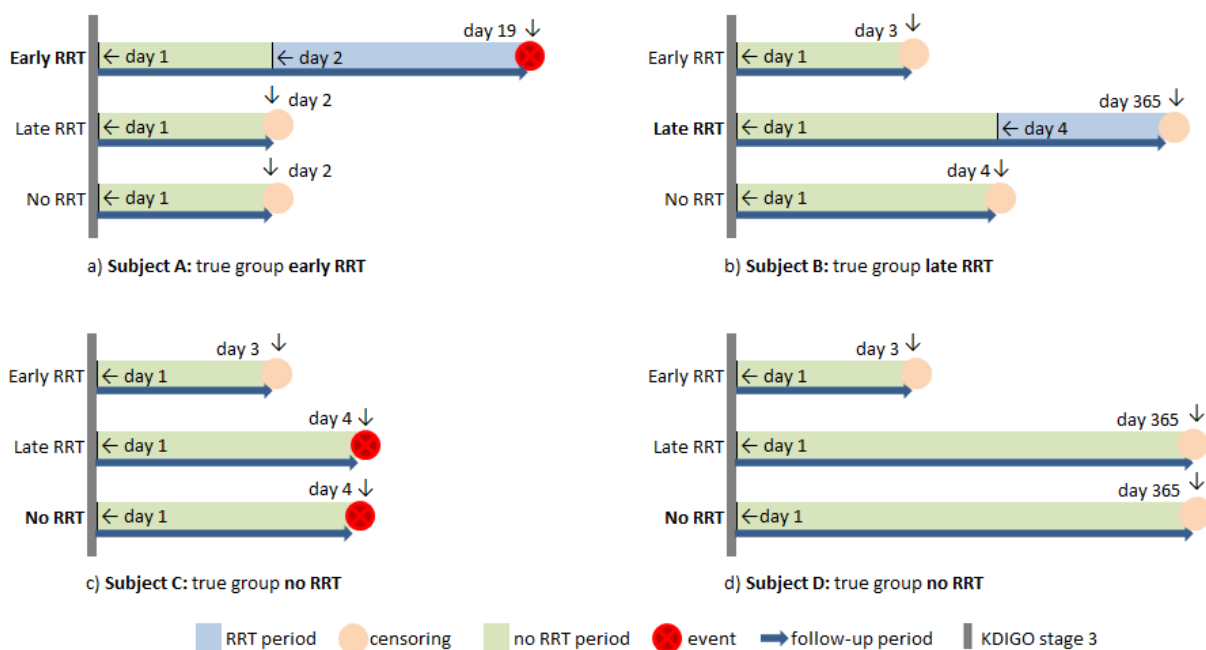


Figure 3: Follow-up of subjects from different RRT groups in the expanded risk sets

See section 2.1.4 for a detailed explanation. RRT, renal replacement therapy; KDIGO stage 3 is the baseline time from which survival is measured.

A survival analysis based on the ERS does not suffer from lead-time bias or immortal-time bias.¹⁶ Lead-time bias is prevented by starting the follow-up time for survival from KDIGO stage 3 for all subjects. Immortal-time bias is prevented by not defining the RRT group using the follow-up time²¹ and also by not excluding subjects who die or become censored before they start RRT.¹⁶

2.2 TIMING OF RRT STUDY DATA

2.2.1 Study Population

This retrospective cohort study used the High-Density Intensive Care (HiDenIC-8) database, which includes data on a source population of 45,568 adult patients admitted to one of 8 ICUs (i.e. medical, cardiac, transplant, surgical, neurological and trauma) within a single academic medical center (University of Pittsburgh Medical Center (UPMC), Pittsburgh, PA) during an 8-year period (July 2000 through October 2008). HiDenIC-8 data was obtained from several computerized databases and deidentified using an honest broker as previously described.²²

For this study, we selected a population of patients that reached KDIGO stage 3 during their hospital stay and: 1) had no prior history of hemodialysis or renal transplant; 2) their known baseline creatinine was < 4 ; 3) had no liver transplant during hospitalization; 4) had no history of heart failure. We were able to identify 4781 such patients. Furthermore, in order to ensure that the population selected had a comparable risk of being started on RRT we applied the following exclusions: 1) RRT started within 24 hours from ICU admission (n=199); 2) in the group of patients that did not receive RRT an increase in serum creatinine (sCr) within 48 hours from KDIGO stage 3 was not observed (n=1860); 3) KDIGO stage was classified on urine output (UO) only and the previous rule could not be determined (n=121); 4) no data on risk factors of interest (n=676). The remaining 1925 patients formed our study population (**Figure 4**). This study was conducted in accordance with institutional review board guidelines and approval.

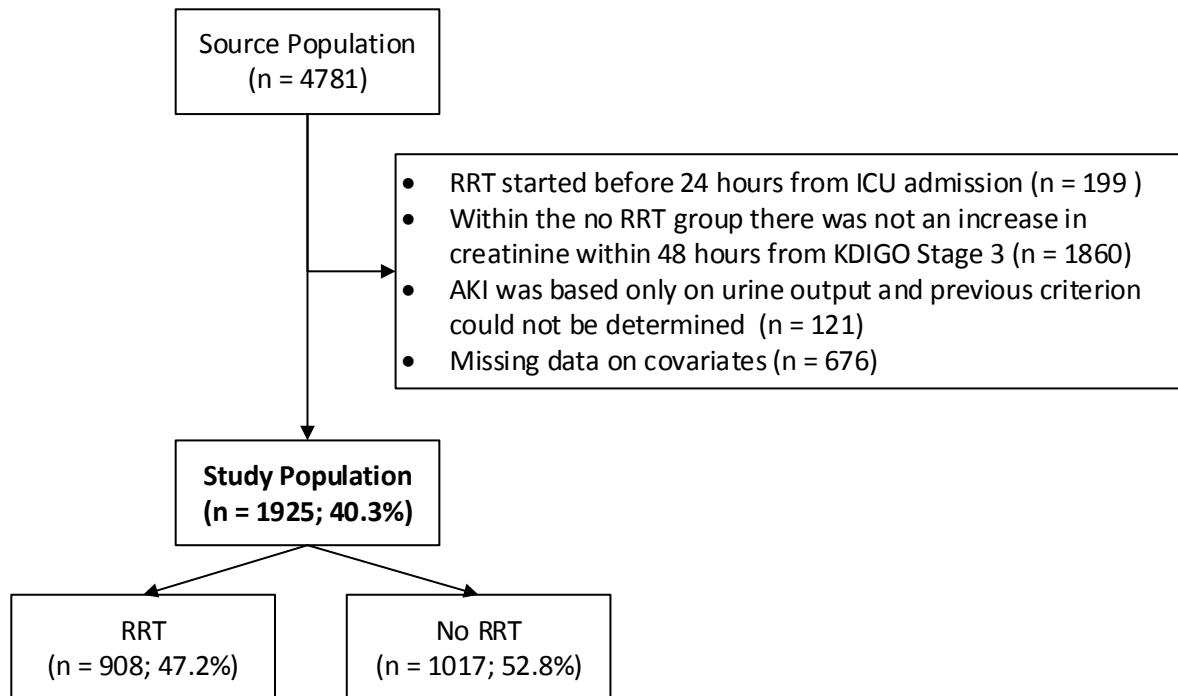


Figure 4: Study population

RRT, renal replacement therapy; ICU, intensive care unit; KDIGO, Kidney Disease Improving Global Outcomes; AKI, acute kidney injury.

2.2.2 Data Collection

Data variables included demographic data, comorbid conditions, and indications for RRT. Demographic data consisted of age, sex and race. History of cardiac disease, chronic renal disease, diabetes and liver transplant were considered. Reference creatinine was derived as previously described.²² Admission type (medical versus surgical) was based on the diagnosis related group at hospital admission. Biochemistry data such as sCr, fraction of inspired oxygen (FiO₂), serum potassium (sK⁺), serum bicarbonate (HCO₃) and serum blood urea nitrogen (BUN) were extracted. Severity scores included Glasgow Coma Scale (GCS), APS-III score and severity of hypotension. Fluids infused, weight adjusted urine, suspected sepsis, use of vasopressors and

mechanical ventilation support were also considered. All variables were measured in the 24 hours following ICU admission and their definitions have been previously described.²²

Patients were classified according to their maximum KDIGO criteria met during hospitalization using sCr and UO criteria.² If multiple episodes of KDIGO stage 3 occurred we only considered the first one as our entry criteria.

2.2.3 Timing of RRT

The first instance of intermittent hemodialysis or continuous RRT was considered as the time of initiation of RRT. Early RRT was defined based on the number of calendar days from KDIGO stage 3 to initiation of RRT. The definition varied from 1 to 7 days. Patients that were started on RRT later than the cut off day or those that were never started on RRT during their hospitalization were used as the control group.

2.2.4 Outcome Assessment

The primary end point of this study was 1-year mortality from KDIGO stage 3. The survival period was calculated from KDIGO stage 3 to mortality (in non-survivors) or censored at 1-year (in survivors).

2.2.5 Statistical Analyses

Statistical analyses were performed using STATA (version SE 11.2), with statistical significance set at p-value <0.05. Mahalanobis matching was done in SAS (version 9.3). Graphs were created in Microsoft Excel 2010 unless otherwise specified. Categorical variables were summarized as frequency (percentage) and continuous variables were summarized as median±interquartile range. For categorical variables the Pearson Chi-square asymptotic test was used and for continuous variables the Kruskal Wallis test was used. First, to determine the propensity for early RRT we ran multivariable logistic models with all risk factors from **Table 2**. All variables were retained in the model regardless of significance level. For a sensitivity analysis, however, backward stepwise selection was used with the probability-to-enter set at 0.05 and probability-to-remove set at 0.1. In this procedure, removal testing was based on the probability of the likelihood-ratio statistic based on conditional parameter estimates. Second, matches from the pool of late/no RRT patients were chosen without replacement using a 1-1 nearest neighbor Mahalanobis distance algorithm. The propensity for early RRT from the logistic regression along with the reference creatinine, FiO₂, sK⁺, fluids and weight adjusted urine were used in calculating the Mahalanobis distance. Covariate balance between groups was checked by plotting the chi-square statistics from the unmatched and matched populations. Third, the ERS method was applied to the unmatched and to the matched populations. We used Cox proportional hazards regression adjusted for age to test the differences in the hazard of mortality at 1-year between late RRT versus early RRT and between no RRT versus early RRT. All steps were applied to each of the 7 populations from **Table 3**.

3.0 RESULTS

3.1 BASELINE CHARACTERISTICS

Of the 1925 patients meeting the inclusion criteria, 47.2% were started on RRT after reaching KDIGO stage 3. Baseline characteristics for the RRT and no RRT groups are shown in **Table 2**. As expected, younger patients, liver transplants, multiple comorbidities, higher reference creatinine, surgical admission, higher FiO₂, azotemia (BUN ≥ 100), worse APS-III scores, more fluids, lower weight adjusted urine, suspected sepsis and use of vasopressors were more common in the RRT group (p-values 0.03 to <0.001). There was no difference in hyperkalemia (sK⁺ > 5 meq/L) (p-value 0.12).

Table 2: Patient characteristics by in-hospital RRT

Characteristics	RRT	No RRT	All	P-value*
	(N = 908)	(N = 1017)	(N = 1925)	
Age	60 (49-70)	64 (51-76)	62 (50-73)	<0.001
Males	526 (57.9)	544 (53.5)	1,070 (55.6)	0.05
Race				
White	682 (75.1)	750 (73.7)	1,432 (74.4)	0.79
Black	58 (6.4)	68 (6.7)	126 (6.5)	
Other	168 (18.5)	199 (19.6)	367 (19.1)	
Comorbid condition				
Cardiac disease	35 (3.9)	47 (4.6)	82 (4.3)	0.41
Chronic renal disease	53 (5.8)	46 (4.5)	99 (5.1)	0.19
Diabetes	137 (15.1)	158 (15.5)	295 (15.3)	0.79
Liver transplant	59 (6.5)	39 (3.8)	98 (5.1)	0.008
Multiple comorbidities	396 (43.6)	394 (38.7)	790 (41)	0.03
Reference creatinine, mg/dl	1 (0.8-1.2)	0.9 (0.8-1.1)	1 (0.8-1.1)	<0.001
Surgical admission	577 (63.5)	568 (55.9)	1,145 (59.5)	0.001
sCr ≥ 4 meq/L^a	152 (16.7)	154 (15.1)	306 (15.9)	0.34
FiO₂ > 60%^a	258 (28.4)	240 (23.6)	498 (25.9)	0.02
sK⁺ > 5 meq/L^a	188 (20.7)	182 (17.9)	370 (19.2)	0.12
HCO₃ < 18 meq/L^a	82 (9)	73 (7.2)	155 (8.1)	0.14
BUN ≥ 100 mgs/dl^a	44 (4.8)	29 (2.9)	73 (3.8)	0.02
GCS^a				
[3,5]	234 (25.8)	250 (24.6)	484 (25.1)	0.36
[6,10]	281 (30.9)	294 (28.9)	575 (29.9)	
[11,15]	393 (43.3)	473 (46.5)	866 (45)	
APS-III score^a	87 (67-112)	79 (59-103.5)	83 (62-109)	<0.001
Severity of hypotension^{a,b}	1.5 (0-15.1)	0.5 (0-13)	1 (0-14)	0.05
Fluids, L^a	4.4 (2.7-6.9)	3.7 (2.3-6.1)	4 (2.5-6.5)	<0.001
Weight adjusted urine, CCs^a	11.3 (4.5-20.3)	13.1 (7.2-23)	12.3 (5.9-21.4)	<0.001
Suspected sepsis^a	303 (33.4)	219 (21.5)	522 (27.1)	<0.001
Vasopressors^a	430 (47.4)	411 (40.4)	841 (43.7)	0.002
Mechanical ventilation^a	604 (66.5)	649 (63.8)	1,253 (65.1)	0.21

Data presented as n (%) or median (Q1-Q3). RRT, renal replacement therapy; sCr, serum creatinine; FiO₂, fraction of inspired oxygen; sK⁺, serum potassium; HCO₃, serum bicarbonate; BUN, serum blood urea nitrogen; GCS, Glasgow Coma Scale; APS-III, acute physiology score.

*P-value for the comparison of RRT and No RRT. For categorical variables Pearson Chi-square asymptotic 2-sided test was used. For continuous variables Kruskal Wallis Test was used.

^aMeasured within 24 hours following ICU admission.

^bArea under the curve for severity and duration of hypotension.

3.2 MATCHING

3.2.1 Generating Propensity Scores

For this study we used the days from KDIGO stage 3 to RRT initiation [median (interquartile range): 3 (2-7)] to define early RRT. We had no a priori definition for the number of days that should classify patients as early starters. Instead, we looked at various cut off points where early was defined as having started RRT on the same day as KDIGO stage 3 or anywhere up to and including day 7. As seen in **Table 3** under the unmatched analysis (*i.e.* the study population), in population 1 there were 192 patients that started RRT on the same day as KDIGO stage 3, 716 that started RRT anywhere after day 2 and 1017 that were never on RRT for a total of 1925 patients. In population 2, there were 128 more patients that started RRT on day 2, thus the early group had 320 patients while the late group was left with 588.

We used the same logistic regression model to generate the propensity for early RRT for each of the 7 unmatched populations:

$$\text{logit} \{P(Y_i = 1 | X_i)\} = \log \left\{ \frac{P(Y_i = 1 | X_i)}{P(Y_i = 0 | X_i)} \right\} = \beta_0 + \beta_1 X_i,$$

where

$Y = 1$ for early RRT,

$Y = 0$ for the combined late and no RRT groups,

X = vector of all baseline covariates from **Table 2**.

Table 3: Sample size description by early RRT definition and analysis type

Days from KDIGO Stage 3 to RRT Initiation	Unmatched Analysis*			Matched Analysis [^]		ERS Analysis [#] (n)
	Early RRT (n)	Late RRT (n)	No RRT (n)	Late RRT (n)	No RRT (n)	
Population 1: 1	192	716	1017	76	116	1152
Population 2: ≤ 2	320	588	1017	139	181	1920
Population 3: ≤ 3	470	438	1017	161	309	2820
Population 4: ≤ 4	568	340	1017	161	407	3408
Population 5: ≤ 5	636	272	1017	147	489	3816
Population 6: ≤ 6	673	235	1017	150	523	4038
Population 7: ≤ 7	703	205	1017	133	570	4218

* Early RRT + Late RRT = 908 and Early RRT + Late RRT + No RRT = 1925 regardless of population.

[^] Late RRT + No RRT = Early RRT; under each matched population, the numbers in the late and no RRT groups represent the number of matched patients from the available pool of unmatched patients in the late and no RRT groups.

[#] Early RRT + (Late RRT + No RRT) [^] multiplied by 3 will give the ERS (n).

3.2.2 Generating Matched Populations

For each population from **Table 3** each subject i from the early RRT group was matched without replacement to one subject j from the late or no RRT group by using the nearest neighbor Mahalanobis distance algorithm. The propensity for early RRT and the following baseline covariates, identified a prior as being clinical indicators for RRT initiation, were used in calculating all M_d 's for each (i, j) pair: reference creatinine, FiO₂, sK⁺, fluids and the weight adjusted urine. **Table 3**, under the matched analysis, gives the number of matched patients and the group from which they originated. For example, in population 1 all 192 patients from the early RRT group were matched to 76 out of 716 and 116 out of 1017 patients from the available late and no RRT groups respectively.

For each population, we checked the distribution and common support of the propensity scores between the early RRT group and the matched and unmatched late and no RRT groups. **Figure 5** shows adequate overlap in the propensity scores between the early RRT

group and the matched late and no RRT groups. In other words, each late and no RRT patient had a good match on the propensity score to an early RRT patient. Population 7 not shown but the results were similar.

We also checked the balance before and after matching on all baseline risk factors that were used to generate the propensity for early RRT. **Figure 6** displays a plot of the chi-square statistics for the test of difference in risk factors between early RRT and the combined late and no RRT groups before matching (red) and after matching (green). For categorical variables we used the Pearson chi-square statistic²³

$$\chi^2 = N \sum_{i=1}^n p_i \left(\frac{O_i / N - p_i}{p_i} \right)^2 \sim \chi_k^2,$$

where

N = the total number of observations,

n = the number of cells compared,

p_i = the proportion of observations of type i ,

O_i = the number of observations of type i ,

χ_k^2 = chi-square distribution with k degrees of freedom.

For continuous variables we used the Kruskal-Wallis statistic²⁴

$$H = \frac{12}{N(N+1)} \sum_{i=1}^c \frac{R_i^2}{n_i} - 3(N+1) \left/ \left\{ 1 - \sum_{j=1}^g (t_j^3 - t_j) / (N^3 - N) \right\} \right. \sim \chi_{c-1}^2,$$

where

c = the number of samples,

n_i = the number of observations in sample i ,

$N = \sum n_i$, the number of observations in all samples combined,

R_i = the sum of the ranks in sample i ,

g = the number of groups with tied observations,

t_j = the number of tied observations in group j ,

χ^2_{c-1} = chi-square distribution with $c-1$ degrees of freedom.

The central tendency line (gray) represents the chi-square value of 3.84 which is analogous to a p-value of 0.05 for a chi-square distribution with 1 degree of freedom. Before matching, across all populations, there was imbalance in most risk factors as represented by the red symbols associated with high chi-square values. However, after matching, the imbalance between risk factors was corrected as represented by the green symbols with very low chi-square values (Figure 6).

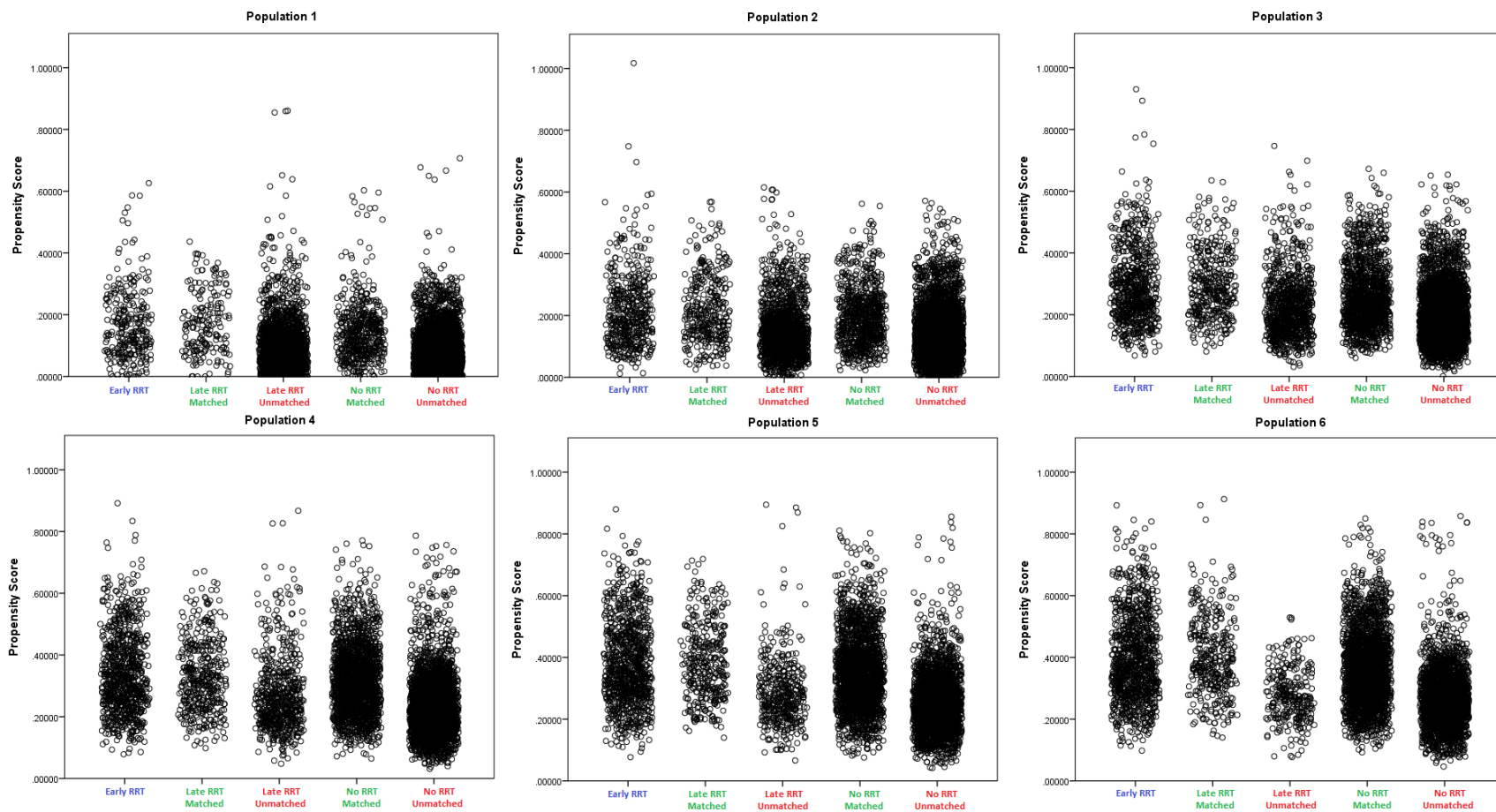


Figure 5: Jitter plot of propensity scores in the matched and unmatched groups

Matches chosen without replacement using a 1-1 nearest neighbor Mahalanobis distance algorithm on the propensity for early RRT, reference creatinine, FiO_2 , sk^+ , fluids and weight adjusted urine.

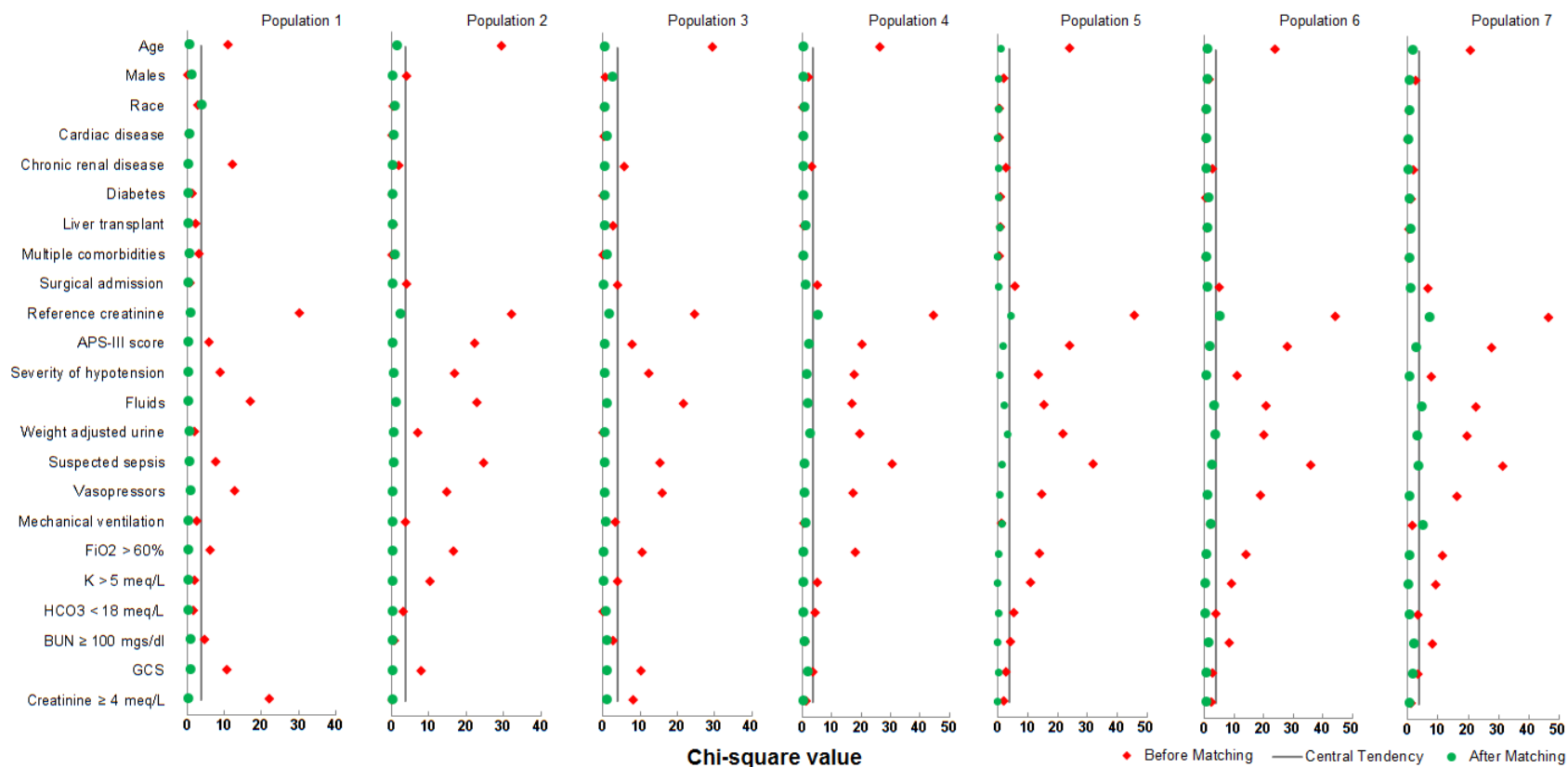


Figure 6: Covariate balance before and after matching

All tests had 1 degree of freedom ($\chi^2_{1,0.05} = 3.84$) except for Race and GCS which had 2 ($\chi^2_{2,0.05} = 5.99$). For a detailed explanation see section 3.2.2.

3.2.3 Generating Expanded Risk Sets Populations

For the ERS analysis, patients from each matched population were allowed to have different contributions of follow-up times and events in all three risk sets: early RRT, late RRT, and no RRT. As a result, the ERS analysis population will have 3 times more subjects than the corresponding matched population. For example, population 1 had $192 \times 2 = 384$ subjects in the matched analysis and $384 \times 3 = 1152$ in the ERS analysis (**Table 3**).

3.3 SURVIVAL

3.3.1 Before Matching

In the unmatched populations, unadjusted analyses showed no difference in hospital mortality between early and late starters except for population 5, where late starters had higher hospital mortality than early starters: 62.9% vs 54.1% (p-value 0.01) (**Table 4**). The unadjusted 1-year mortality was significantly worse in late than in early starters as the definition for early RRT changed from ≤ 3 to ≤ 7 days (**Table 4**). There was no apparent benefit in initiating RRT earlier in the course of the disease progression since the 1-year mortality in the early group only slightly increased from 66.9% in population 2 to 68.8% in population 7. However, these results suffer from selection bias and immortal-time bias.

Table 4: Unadjusted outcomes before matching

Early RRT	Hospital Mortality (%)				1-year Mortality (%)			
	Early RRT	Late RRT	No RRT	P-value*	Early RRT	Late RRT	No RRT	P-value*
Population 1	59.4	56		0.4	69.8	70.9		0.75
Population 2	55.6	57.3		0.62	66.9	72.8		0.06
Population 3	55.3	58.2		0.38	67.2	74.4		0.02
Population 4	54.2	60.9	41.3	0.05	67.8	75.6	57.7	0.01
Population 5	54.1	62.9		0.01	67.9	77.2		0.005
Population 6	54.8	62.1		0.05	68.8	76.2		0.03
Population 7	55.2	62		0.09	68.8	77.1		0.02

The sample size for each population and RRT group is the same as in **Table 3** under the unmatched analysis.

*P-value for the comparison between early RRT and late RRT groups only.

All 3-way comparisons had p-value < 0.001.

In the unmatched populations, after adjusting for age and taking into account the time to death there was no difference in the hazard of mortality at 1-year between early and late starters (**Figure 7a**). Even though the hazard ratio (HR) in population 1 (HR (95%CI): 0.86 (0.71-1.04)) seemed to favor late initiation the confidence interval contained 1. As definitions for early RRT changed the HRs got very close to 1 and there was no clear signal of an optimal time for initiation. However, these results suffer from both selection bias and immortal-time bias.

Next, we removed the artificial survival advantage given to late starters by applying the ERS method to the unmatched populations. The sample size for each population was 5775 (1925*3). In this analysis, there seems to be an advantage in delaying RRT since the rate of mortality for the late RRT group decreased from 14% in population 2 to 25% in population 7 (**Figure 7b**). Selection bias has not been addressed in the ERS analysis on the unmatched populations.

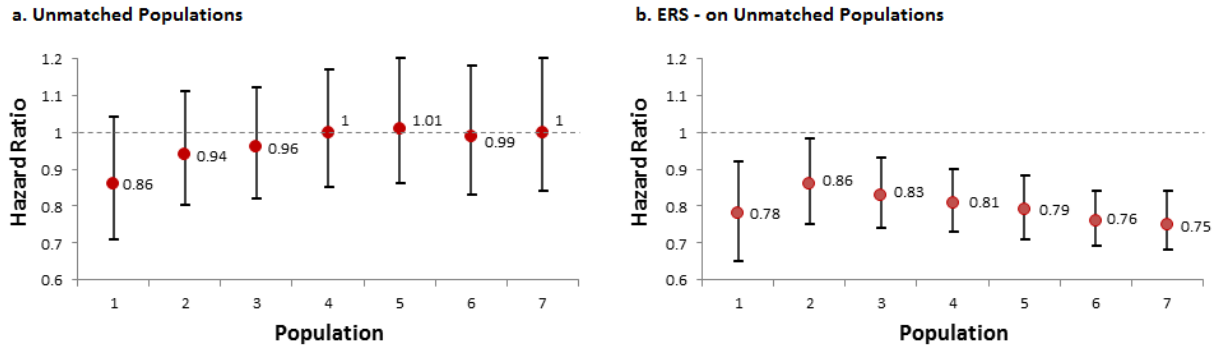


Figure 7: Age adjusted survival at 1-year before matching

Red dots represent the hazard ratios (HR) for late RRT vs early RRT; black lines represent 95% confidence intervals for HR; dashed lines represent a HR of 1; ERS, expanded risk sets. Sample size for each population in (a) is 1925. Sample size for each population in (b) is $1925 \times 3 = 5775$.

3.3.2 After Matching

In the matched analysis, after controlling for age, there was no difference in the hazard of mortality at 1-year associated with timing of RRT initiation. Even though in population 1 the HR of 0.84 seemed to favor late initiation the 95% confidence interval (0.61-1.16) contained 1 (**Table 5**). As definitions for early RRT changed the HRs got very close to 1 and there was no clear signal of an optimal time for initiation. Except for populations 2 and 5 patients that were never started on RRT seemed to have a decreased hazard of 1-year mortality when compared to early starters (**Table 5**). However, these results still suffer from immortal-time bias.

In the ERS analysis the HRs were lower for both late and no RRT groups but their magnitude did not change with the varying definitions for early RRT. In population 1 the rate of mortality decreased by 25% in the late RRT group and by 44% in the no RRT group. When the definition of early RRT was changed, the rates of mortality in the late RRT group only slightly varied from 24% in population 3 to 27% in population 7 with no significant decrease in mortality for population 2. Similar to the results seen in **Figure 7b** which only suffered from selection

bias, there seems to be an advantage in delaying RRT once both selection bias and immortal-time bias were removed through the use of the ERS analysis.

Table 5: Survival at 1-year in the matched and expanded risk sets populations

Early RRT	Matched Analysis			ERS Analysis		
	N	HR* (95%CI)		N	HR* (95%CI)	
		Late vs Early	No RRT vs Early		Late vs Early	No RRT vs Early
Population 1	384	0.84 (0.61-1.16)	0.71 (0.53-0.96)	1152	0.75 (0.59-0.96)	0.56 (0.42-0.76)
Population 2	640	0.93 (0.73-1.18)	0.85 (0.67-1.07)	1920	0.85 (0.71-1.03)	0.63 (0.5-0.79)
Population 3	940	0.88 (0.71-1.09)	0.78 (0.65-0.94)	2820	0.76 (0.65-0.89)	0.59 (0.49-0.71)
Population 4	1136	0.99 (0.8-1.22)	0.8 (0.68-0.94)	3408	0.78 (0.67-0.88)	0.6 (0.52-0.71)
Population 5	1272	0.92 (0.74-1.13)	0.87 (0.75-1.01)	3816	0.77 (0.67-0.87)	0.64 (0.55-0.73)
Population 6	1346	0.93 (0.75-1.15)	0.81 (0.7-0.94)	4038	0.73 (0.65-0.83)	0.6 (0.53-0.69)
Population 7	1406	0.96 (0.77-1.19)	0.83 (0.72-0.95)	4218	0.73 (0.65-0.83)	0.61 (0.54-0.7)

RRT, renal replacement therapy; HR, hazard ratio; CI, confidence interval; ERS, expanded risk sets; *Age adjusted.

For a sensitivity analysis we determined the propensity for early RRT by using multivariable logistic regression with backward stepwise variable selection as described under section 2.2.5. We then followed all other steps for matching and for creating ERS populations. Changing the method for selecting the propensity for early RRT did not modify our overall results. There still seems to be an advantage in delaying RRT once both selection bias and immortal-time bias were removed through the use of the ERS analysis (**Figure 8a**).

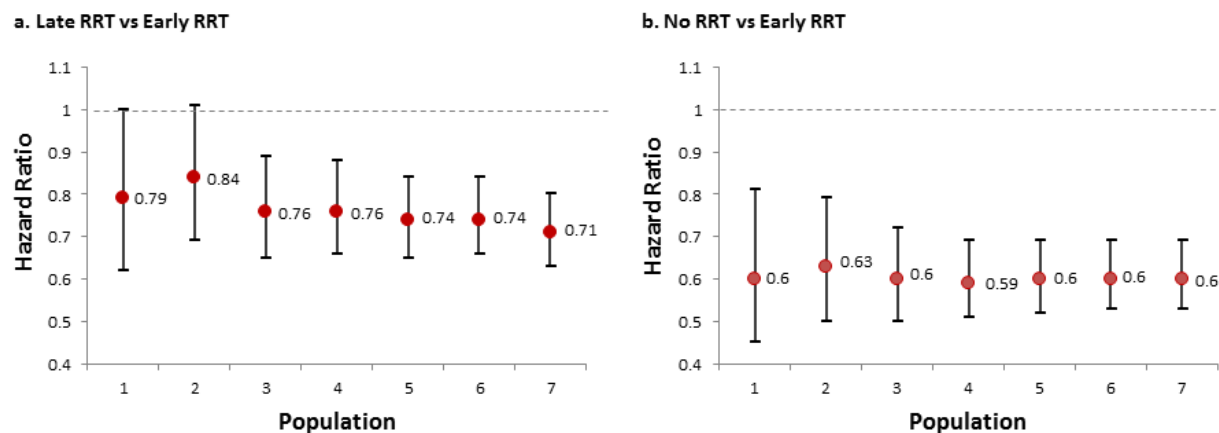


Figure 8: Age adjusted survival at 1-year after matching

Red dots represent the hazard ratios (HR) for late RRT vs early RRT (a) and for no RRT vs early RRT (b); black lines represent 95% confidence intervals for HR; dashed lines represent a HR of 1; ERS, expanded risk sets. Populations were generated from ERS models based on matched populations where the propensity for early RRT was modeled with a backward stepwise selection method. The sample size for each population is the same as in the ERS analysis from **Table 3** but the sample size for the true late and true no RRT groups is slightly different.

4.0 DISCUSSION

In this retrospective cohort study, using a large adult population admitted to one of 8 ICUs within a single academic medical center over an 8-year period, we examined the association between timing of initiation of RRT and 1-year mortality. Current studies in this domain suffer from either selection bias, lead-time bias, and/or immortal-time bias. To our knowledge, in the renal literature, there is only one other study that addresses both lead-time bias and immortal-time bias.¹⁶ However, there are several major differences between our studies. First, we defined the groups early RRT, late RRT, and no RRT based on the number of days from KDIGO stage 3 to RRT initiation (**Table 3**) and not on changes in estimated glomerular filtration rate. Second, we varied our definition for early RRT from 1 to 7 days to find the optimal time for RRT initiation. Third, before expanding the risk sets we dealt with selection bias. In conclusion, by using a new definition for the timing of initiation and statistical techniques based on propensity scores, Mahalanobis matching, and ERS analyses we have addressed three of the existing biases in the current literature.

In randomized control trials, on average, patients are similar on all baseline characteristics. Hence, any significant differences between groups in the outcome event can be attributed to the intervention.²⁵ However, in observational studies, the assignment of patients into the treatment and control groups is typically not random. Differences in the outcome may

not necessarily be due to the treatment effect but rather to the differential distributions of other prognostic factors, known and unknown, which are associated with both the outcome and the decision to treat.^{26–28} We believe that our study contains patients in the late RRT and no RRT groups in which due to their baseline risk factors and day-to-day disease progression the decision to initiate RRT was never a choice. This is supported by the divergent distributions of risk factors between the early RRT and late/no RRT groups (**Figure 6**). Even though with the available risk factors we cannot address the daily changes in a patient's risk to be initiated on RRT, we believe that the spectrum of their baseline risk factors is a good indication for the decisions made. Although no method can be trusted to remove hidden selection bias, it is important to minimize differences between known risk factors.²⁶ Thus, as suggested in literature, the use of a propensity score along with Mahalanobis metric matching can eliminate hidden selection bias.^{18,29,30}

Lead-time bias occurs in observational studies when follow-up time for survival is counted from the exposure time rather than from enrollment. Patients exposed to the treatment earlier in the course of their disease development get an artificial survival time over those that are exposed later. Even though it has been acknowledged as a limitation by several authors^{9,31,32} dating to 2001, studies affected by this bias have been published in 2010³³ and as recently as 2014³⁴. In our study, survival time was counted from the time patients reached the same severity in their renal dysfunction, namely KDIGO stage 3.

Immortal time in epidemiology refers to a period of cohort follow-up or observation time, during which death cannot occur.³⁵ Depending on the methodology used, if immortal time is not correctly accounted for, estimated treatment effects can be substantially biased.³⁶

Pharmacoepidemiology research has shown that for time-based, event-based, and exposure-based cohort definitions, the bias in the rate ratio resulting from misclassified or excluded immortal time increases proportionately to the duration of immortal time.^{37,38} A review of 127 studies published in highly-cited medical journals found that immortal time was not handled properly in 52 of them.³⁹

Immortal time bias can occur in observational studies in one of two ways. The first is through misclassification of immortal time as a part of the follow-up time for survival. As a result, patients with longer immortal time periods have an artificially inflated survival advantage when compared to patients with shorter immortal time periods. The second is through exclusion of immortal time from the analysis and then starting the follow-up time for survival at the exposure time. Immortal time bias through exclusion differs from lead-time bias in that it occurs when patients who were never exposed to the treatment of interest are now included in the survival analysis. In this case, the exposed and unexposed patients are not comparable because their follow-up times start at different stages in the development of their disease. In the ERS analysis the RRT group assignment is made at baseline (*i.e.* KDIGO stage 3) thus the time-varying nature of the treatment is removed and patients no longer have immortal time. For example, within the early RRT group, subjects that initiated RRT or those that died or become censored before initiating RRT all contribute their observed follow-up days for survival to the same treatment regime – early RRT, thus within group immortal-time bias has been removed.

There are important limitations to our study. First, because this was an observational study our results are subject to unmeasured confounding (*i.e.* hidden bias) and causation

cannot be established. Even though we used a propensity score to minimize selection bias, we only used patient specific parameters to address this issue and there are immeasurable physician beliefs and logistical and/or operational issues that impact RRT initiation that we did not account for.¹³ Second, even though we had a large sample size, patients are all from a single medical center which makes it difficult to assess the generalizability of our results. However, we had access to patient data from multiple ICUs (*i.e.* medical, cardiac, transplant, surgical, neurologic, and trauma) which increases our confidence that the results are not unique to this medical center. Third, by using the ERS method we introduced nonrandom censoring and patients censored at time t will have worse prognosis than uncensored patients.¹⁶ In the future, we plan to use inverse probability weighting to adjust for any new selection bias introduced by the nonrandom censoring.^{16,40}

In this study, we found that selection bias did not have a big impact on the estimated hazard of mortality but immortal-time bias drastically affected the conclusions drawn. When comparing results from unmatched and matched analyses that still suffered from immortal-time bias we concluded that there was no clear signal of an optimal time to initiate RRT (**Figure 7a** versus **Table 5** - matched analyses). However, after removing immortal-time bias, regardless of selection bias, there seemed to be an advantage in delaying RRT (**Figure 7b** and **Table 5** - ERS analysis).

In conclusion, the optimal time to start dialysis is still uncertain but we believe that building upon our methods and those used by Sjolander et al.¹⁶ will aid future researchers in better analyzing observational data and providing less biased estimates.

BIBLIOGRAPHY

1. Hoste EAJ, Clermont G, Kersten A, et al. RIFLE criteria for acute kidney injury are associated with hospital mortality in critically ill patients: a cohort analysis. *Crit Care*. 2006;10:R73. doi:10.1186/cc4915.
2. KDIGO AKIWG. KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney Int Suppl*. 2012;2:1–138. doi:10.1038/kisup.2012.1.
3. Uchino S, Kellum JA, Bellomo R, et al. Acute renal failure in critically ill patients: a multinational, multicenter study. *JAMA*. 2005;294:813–818. doi:10.4045/tidsskr.10.34549.
4. Van Biesen W, Vanholder R, Lameire N. Dialysis strategies in critically ill acute renal failure patients. *Curr Opin Crit Care*. 2003;9:491–495. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14639068>.
5. Gettings LG, Reynolds HN, Scalea T. Outcome in post-traumatic acute renal failure when continuous renal replacement therapy is applied early vs. late. *Intensive Care Med*. 1999;25:805–813. doi:10.1007/s001340050956.
6. Demirkilic U, Kuralay E, Yenicesu M, et al. Timing of replacement therapy for acute renal failure after cardiac surgery. *J Card Surg*. 2004;19:17–20. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15108784.
7. Elahi MM, Yann Lim M, Joseph RN, Dhannapuneni RR V, Spyt TJ. Early hemofiltration improves survival in post-cardiotomy patients with acute renal failure. In: *European Journal of Cardio-thoracic Surgery*. Vol 26.; 2004:1027–1031. doi:10.1016/j.ejcts.2004.07.039.
8. Liu KD, Himmelfarb J, Paganini E, et al. Timing of initiation of dialysis in critically ill patients with acute kidney injury. *Clin J Am Soc Nephrol*. 2006;1(5):915–919. doi:10.2215/CJN.01430406.

9. Korevaar JC, Jansen MAM, Dekker FW, et al. When to initiate dialysis: Effect of proposed US guidelines on survival. *Lancet*. 2001;358:1046–1050. doi:10.1016/S0140-6736(01)06180-3.
10. Bouman CS, Oudemans-Van Straaten HM, Tijssen JG, Zandstra DF, Kesecioglu J. Effects of early high-volume continuous venovenous hemofiltration on survival and recovery of renal function in intensive care patients with acute renal failure: a prospective, randomized trial. *Crit Care Med*. 2002;30:2205–2211. doi:10.1097/01.CCM.0000030444.21921.EF.
11. Shiao CC, Ko WJ, Wu VC, et al. U-Curve Association between Timing of Renal Replacement Therapy Initiation and In-Hospital Mortality in Postoperative Acute Kidney Injury. *PLoS One*. 2012;7. doi:10.1371/journal.pone.0042952.
12. NKF-DOQI clinical practice guidelines for peritoneal dialysis adequacy. National Kidney Foundation. *Am J Kidney Dis*. 1997;30:S67–S136. doi:10.1016/S0272-6386(97)70028-3.
13. Gibney N, Hoste E, Burdmann EA, et al. Timing of initiation and discontinuation of renal replacement therapy in AKI: Unanswered key questions. *Clin J Am Soc Nephrol*. 2008;3:876–880. doi:10.2215/CJN.04871107.
14. Palevsky PM. Clinical review: timing and dose of continuous renal replacement therapy in acute kidney injury. *Crit Care*. 2007;11:232. doi:10.1186/cc6121.
15. Rosenbaum PR, Rubin DB. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *J Am Stat Assoc*. 1984;79:516–524. doi:10.2307/2288398.
16. Sjölander A, Nyrén O, Bellocco R, Evans M. Comparing different strategies for timing of dialysis initiation through inverse probability weighting. *Am J Epidemiol*. 2011;174:1204–1210. doi:10.1093/aje/kwr249.
17. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55. doi:10.1093/biomet/70.1.41.
18. Rosenbaum PR, Rubin DB. Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *Am Stat*. 1985;39:33–38. doi:10.1080/00031305.1985.10479383.
19. De Maesschalck R, Jouan-Rimbaud D, Massart DLL. The Mahalanobis distance. *Chemom Intell Lab Syst*. 2000;50:1–18. doi:10.1016/S0169-7439(99)00047-7.
20. Cox DR. Regression models and life tables. *J R Stat Soc Ser B*. 1972;34(2):187–220.

21. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ*. 2010;340:b5087. doi:10.1136/bmj.b5087.
22. Sileanu FE, Murugan R, Lucko N, et al. AKI in Low-Risk versus High-Risk Patients in Intensive Care. *Clin J Am Soc Nephrol*. 2014;1–10. doi:10.2215/CJN.03200314.
23. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos Mag Ser*. 1900;5(50):157–175.
24. Kruskal WH, Wallis WA. Use of Ranks in One-Criterion Variance Analysis. *J Am Stat Assoc*. 1952;47:583–621. doi:10.1080/01621459.1952.10483441.
25. Stolberg HO, Norman G, Trop I. Randomized controlled trials. *AJR Am J Roentgenol*. 2004;183(6):1539–1544. doi:10.2214/ajr.183.6.01831539.
26. Cochran WG, Chambers SP. The Planning of Observational Studies of Human Populations. *J R Stat Soc Ser A*. 1965;128(2):234–266 CR – Copyright © 1965 Royal Statisti. doi:10.2307/2344179.
27. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688–701. doi:10.1037/h0037350.
28. Rosenbaum PR. Observational Study. *Encycl Stat Behav Sci*. 2005;3:1451–1462. doi:10.1002/0470013192.bsa454.
29. Rubin DB. Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Serv Outcomes Res Methodol*. 2002;2:169–188.
30. Stuart E a. Matching Methods for Causal Inference: A Review and a Look Forward. 2010;25(1):1–21. doi:10.1214/09-STS313.
31. Traynor JP, Simpson K, Geddes CC, Deighan CJ, Fox JG. Early initiation of dialysis fails to prolong survival in patients with end-stage renal failure. *J Am Soc Nephrol*. 2002;13:2125–2132. doi:10.1097/01.ASN.0000025294.40179.E8.
32. Lameire N, Biesen W Van, Vanholder R. Initiation of dialysis--is the problem solved by NECOSAD? *Nephrol Dial Transplant*. 2002;17:1550–1552.
33. Wright S, Klausner D, Baird B, et al. Timing of dialysis initiation and survival in ESRD. *Clin J Am Soc Nephrol*. 2010;5:1828–1835. doi:10.2215/CJN.06230909.

34. Lee J, An JN, Hwang JH, et al. Effect of Dialysis Initiation Timing on Clinical Outcomes: A Propensity-Matched Analysis of a Prospective Cohort Study in Korea. *PLoS One*. 2014;9(8):e105532. doi:10.1371/journal.pone.0105532.
35. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
36. Liu J, Weinhandl ED, Gilbertson DT, Collins AJ, St Peter WL. Issues regarding “immortal time” in the analysis of the treatment effects in observational studies. *Kidney Int*. 2012;81:341–350. doi:10.1038/ki.2011.388.
37. Suissa S. Effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease: Immortal time bias in observational studies. *Am J Respir Crit Care Med*. 2003;168:49–53. doi:10.1164/rccm.200210-1231OC.
38. Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol*. 2008;167:492–499. doi:10.1093/aje/kwm324.
39. Van Walraven C, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. *J Clin Epidemiol*. 2004;57:672–682. doi:10.1016/j.jclinepi.2003.12.008.
40. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol*. 2006;98(3):237–242. doi:10.1111/j.1742-7843.2006.pto_329.x.